Clinical trial discrimination of physical function instruments for psoriatic arthritis: a systematic review

Ying Ying Leung[1], Richard Holland[2] Ashish Methew[3], Chris Lindsay[4], Niti Goel[5], Alexis Odgie[6], Ana-Maria Orbai[7], Pil Hojgaard[8], Jeffrey Chau[9], Laura C Coates[10], Vibeke Strand[11], Dafna D Gladman[12], Robin Christensen[13], William Tillett[14], Philip Mease[15]

1. Singapore General Hospital, Duke-NUS Medical School, Singapore
2. Concord Repatriation General Hospital, Sydney, Australia
3. Xxx
4. Patient research partner
5. Patient Research Partner, Adjunct Assistant Professor, Duke University School of Medicine, Durham, North Carolina, USA
6. Medicine and Epidemiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA
7. Division of Rheumatology, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA
8. Copenhagen Lupus and Vasculitis Clinic, Center for Rheumatology and Spine Diseases, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark.
8.
9. Patient Research Partner
10. National Institute for Health Research Clinician Scientist, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, United Kingdom
11. Division of Immunology/Rheumatology, Stanford University School of Medicine, Palo Alto, California, USA

12. Medicine, University of Toronto, Senior Scientist, Krembil Research Institute, Director, Psoriatic Arthritis Program, University Health Network, Toronto Western Hospital, Toronto, Ontario, Canada
13. Xx
14. Royal National Hospital for Rheumatic Diseases, University of Bath, Bath, United Kingdom
15. Rheumatology Research, Swedish Medical Center and University of Washington School of Medicine, Seattle, Washington, USA

Correspondence to: Ying-Ying Leung, MD; Department of Rheumatology and Immunology, Singapore General Hospital, The Academia, level 4, 20 College Road, Singapore 169856, Contact No.: +65 63265276, Fax no.: +65 62203321, E-mail: katyccc@hotmail.com

Abstract word count:

Manuscript word count:

Figures/Tables:

Key words:

Abstract

INTRODUCTION

Psoriatic arthritis (PsA) is a chronic inflammatory disease involving peripheral arthritis, enthesitis, dactylitis, spondylitis, psoriasis and nail disease (1). It has a profound impact on patients' physical, psychological and social well-being. The inflammation of axial and peripheral joints as well as entheses causes pain and possibly joint erosion and destruction (2). Both the acute inflammation and joint damage from PsA causes loss of physical function and disability (3). Physical function is a key concern from numerous qualitative studies among PsA patients (4, 5), and it is recognized as one of the core domains to be measured in every randomized controlled trial (RCTs) and longitudinal observational study (6).

There have been several instruments in the form of patient-reported outcome measures (PROMs) that assess physical function in PsA (7). The most commonly used PROMs for the physical function domain (PF-PROM) in RCTs for PsA have been the Health Assessment Questionnaire Disability Index (HAQ-DI) (8) and the physical functioning domain within the Medical Outcomes Study 36-item Short Form Survey (SF-36 PF) (9). A few additional ones have also been validated and evaluated for use in PsA (10). The discriminative performance of these PF-PROMs in RCTs have not been evaluated systematically. In this systematic review, we aimed to evaluate the clinical trial discrimination properties of PF-PROMs in RCTs of PsA. The data derived from this study contribute to the concerted effort of the Group for Research And Assessment of Psoriasis and Psoriatic Arthritis (GRAPPA) – Outcome Measures in Rheumatology (OMERACT) initiative to standardize an outcome measurement set for PsA (11).

METHODS

The protocol of this systematic review was registered with PROSPERO prior to initiation (CRD42019129557). The report of this systematic review adheres to the Preferred Reporting Items for Systematic reviews and Meta-Analysis (PRISMA statement) (12).

Literature search and eligibility of articles

We performed a database search through 21 March 2019 using PubMed and Scopus. The search aimed to identify all original full text articles of RCTs conducted in PsA with published data in the English language. The detail of search terms is summarized in supplementary table 1. We included articles on RCTs conducted in PsA only. If the RCT was conducted in a mixed arthritis population, we only included those with separate subgroup analyses in PsA. We excluded RCTs that were not double blinded in study design.

For the purpose of deriving responsiveness data for the PF-PROMs, we limited the review to RCTs involving biological (b) disease modifying anti-rheumatic drugs (DMARDs) and targeted synthetic (ts)-DMARDs. For RCTs with multiple publications, only one article from each unique RCT was included unless a publication provided new information, such as subgroup analyses. Trials that did not report data for PF-PROMs were excluded.

Selection of articles

One researcher (YYL) removed duplicates from the searches of two databases. Two researchers (YYL, RH) independently screened the titles, abstracts and full text (if appropriate) for eligibility. Disputes were resolved by consensus of the two researchers and a third researcher if needed. Additional studies identified by co-authors were considered for inclusion. The stepwise eligibility and inclusion of articles are summarized in a flow diagram (Figure 1).

Extraction of study characteristics and description of PROMs

YYL and AO independently extracted data on the characteristics of the studies including gender, duration of illness of participants, proportion of participants taking methotrexate and resistant to tumor necrosis factor inhibitors (TNFi), active interventions, comparisons, primary outcomes and their time points. Researchers worked in groups of two (YYL and RH; YYL and ASM; YYL and NG; YYL and CL) to extract data independently for PF-PROMs at baseline and the time point of assessment of primary outcomes (or end of blinded controlled periods). These assessment time points were chosen with the consideration that they represented responsiveness of the PF-PROM in the control groups, as most RCTs offered cross-over to the active intervention after the primary end point. One researcher (YYL) calculated the effect sizes (ESs) for the PF-PROMs based on published data and online supplementary materials if available. Effect sizes of PF-PROMs comparing active and control groups were evaluated using the following methods wherever data were available (3): 1) ES_1, calculated by change scores divided by standard deviations (SDs) of baseline scores; 2) ES_2, calculated by change scores divided by pooled SDs; and 3) standardized response means (SRMs), calculated by change scores divided by SDs of the change scores.

For papers that provided standard error (SE), SD was calculated using the formula: SD = SE x √sample size (13). For papers that included information on median and interquartile ranges, means (SD) were calculated using an estimation formula suggested by Wan et al (14).

Quality Assessment

We assessed the clinical trial discrimination of PF-PROMs in each article using the OMERACT good method checklist (15). Two researchers in groups (YYL and RH; YYL and ASM; YYL and NG; YYL and CL) appraised each instrument for the following categories using (+) yes, good methods; (+/-) some cautions; or (-) no, not achieved:

1. Was the time interval between testing stated and appropriate?
2. Were there proportions of people expected to change in one or both groups?
3. Were hypotheses described regarding the anticipated mean differences in change scores between subgroups *a priori*?
4. Were the statistical methods adequate for the hypotheses tested?
5. Otherwise good methods? (free of any other important flaws)

The two-researcher groups independently assessed a final decision for each article rated as 1) Green, likely low risk of bias; 2) Amber, some cautions but can be used as evidence and 3) Red, not to be used as evidence. Consensus between the two researchers in the group was sought and disputes were resolved with a third member of the team if necessary.

Appraisal of effect sizes using a priori hypotheses and evidence synthesis

The calculated ESs were appraised with the following *a priori* hypotheses (based on general knowledge of the therapeutic efficacy of various treatments):

1. At the assessment time point (primary outcome or end of blinded controlled period), patients treated with b-DMARDs have significant changes in PF-PROM, whereas patients receiving the control intervention do not (except for alefacept [ALC] and clazakizumab ts-dn[CLZ] where no significant differences were expected).
2. PF-PROM change scores among patients given b-DMARDs are significantly better than those receiving control. PCB.
3. Within individual trials, the PF-PROM ESs are higher for those treated with b-DMARDs compared with PCB, MTX or conventional synthetic (cs)-DMARD, but do not differ significantly with different b-DMARD doses (or with TNFi as active comparators).

4. If data for subgroup analyses are available, ESs of change scores of the PF-PROM are higher in TNF naïve versus TNF exposed subgroups.

We synthesized and summarized the overall evidence to support clinical trial discrimination for each PF-PROM in a Summary of Measurement Properties evidence table (SOMP). The number of articles available for measurement properties and evidence synthesis are presented. For each article, colour coding of Green/Amber/Red indicated the quality assessment and (+), (+/-) or (-) indicated fulfilment, partial fulfilment or not supportive of the hypotheses. Synthesizing evidence from all articles evaluated, an overall rating for clinical trial discrimination of the PF-PROM was given as Green (Good to go) /Amber (Some concerns but good to go) /Red (Stop, do not use this for evidence synthesis), or White for no information available.

RESULTS

Literature search results

Our literature search identified a total of 676 articles. After removal of duplicates, 608 articles remained;. one article was identified through cross referencing check. There were 439 articles excluded for the following reasons: not RCTs (344 articles), open-label trials (10 articles), not in PsA (65 articles), mixed population of arthritis without separate data reporting for PsA (18 articles) and trial protocols (2 articles). There were 170 articles eligible for full text review as RCTs in PsA. We excluded 48 articles as trials for interventions other than b-DMARDs or ts-DMARDs, 67 as secondary analyses of RCTs in PsA, 13 without inclusion of PF-PROM and 9 as pooled data from RCTs. A total of 41 articles were included in the full text review that reported primary data on double-blind RCTs in PsA with b-DMARDs and ts-DMARDs. In total, 33 relevant articles representing 32 RCTs were retained for evidence synthesis and 31 RCTs (75.6%) included data for the physical function domain (Figure 1).

Of the 13 articles reporting 10 RCTs that did include physical function data, one and three were Phase 1b and 2 trials, respectively. Among the six Phase 3 RCTs that did not include physical function data, two were pilot proof of concept studies with molecular and tissue focus, and three were RCTs in psoriasis, not PsA.

In these RCTs physical function was evaluated using four PF-PROMs, including the Health Assessment Questionnaire (HAQ)-Disability Index (HAQ-DI) (8), HAQ-Spondyloarthritis (HAQ-S) (16), Physical Component Summary Score of the SF-36 (SF-36 PCS) (9), and the Physical Functioning domain of the SF-36 (SF-36 PF10) (9). New potential PF-PROMs shortlisted by GRAPPA for further evaluation [paper under review], including the multidimensional HAQ (MDHAQ) and the Patient-Reported Outcomes Measurement Information System (PROMIS)-Short Form Physical Function 10a (PROMIS-PF10a), were not utilized in the RCTs included in this review.

Clinical trial discrimination for PF-PROM

Results were reported for HAQ-DI in 31 articles from 30 unique RCTs . The ESs calculated for both interventional and comparator groups are shown in Table 2. Effect sizes for HAQ-DI could not be calculated from two published articles (detailed in Supplementary Table 6), and data were not used for evidence synthesis. Quality assessment using the OMERACT good method checklist was generally affirmative (Supplementary Table 2). Minor concerns for quality were noted in 10 articles, predominantly resulting from not clearly stating expected change scores in HAQ-DI (although most expected changes were implied), estimating ESs from median/ interquartile ranges, or using percentage changes that raised concerns for introduction of minor errors. There were 29 articles included for evidence synthesis as shown in the SOMP table (Table 6). Results aligned with the *a priori* hypotheses that statistically significant change scores were reported in the active but not control groups

with higher ESs (Table 2). Three articles included subgroup analyses, where one demonstrated higher ESs for HAQ-DI in TNFi naïve compared with TNFi exposed groups (17), while two did not (18, 19). Based on all the evidence, the working group recommended GREEN (+) to support HAQ-DI in RCT discrimination in PsA, indicating a low risk of bias and results aligned with hypotheses.

There was only one article that reported results for HAQ-S (20), showing higher ESs for the ABT-122 and adalimumab groups compared with control which aligned with the hypotheses (Table 3). However, no statistical test was performed for the change scores in different groups. The working group recommended AMBER (+) for HAQ-S (Table 6), recognizing that more data are required for better evidence synthesis.

The SF-36 was included in most RCTs in PsA. However, of the 33 included articles, only 24 reported SF-36 PCS results and four SF-36 PF. Of the articles included for SF-36 PCS evidence synthesis, one did not pass the OMERACT checklist for evidence synthesis. It failed to include adequate data for ES calculations (21); also compared with control groups, statistically significant differences in change scores were evident with the higher dose but not lower dose used in the interventional groups (Table 4). For the 23 eligible articles, most results aligned with the *a priori* hypotheses with minor quality concerns in 7 similar to those stated for HAQ-DI (Supplementary Table 4). The working group recommended GREEN (+) supporting clinical trial discrimination with SF-36 PCS.

Four articles reported SF-36 PF, and two did not have adequate data for ES estimations and were excluded from evidence synthesis (Table 5). Of the remaining 2 articles, both with tofacitinib [ref], higher ESs in the interventional groups were shown compared with control (Table 5). Based on the limited evidence, the working group recommended AMBER (+) for clinical trial discrimination with SF-36 PF (Table 6).

DISCUSSION

In this systematic review, we summarized the clinical trial discrimination of the available PF-PROMs in PsA. Of 41 unique RCTs in PsA with b-DMARD/ ts-DMARDs, 31 (75.6%) reported results of at least one measurement of the physical function domain. This is the first paper that systematically appraised the clinical trial discrimination properties for PF-PROMs. Numerous instruments are available for assessing physical function in PsA (7, 10). However, data for appraisal of clinical trial discrimination was available for only four PF-PROMs (HAQ-DI, HAQ-S, SF-36 PCS and SF-36 PF). The majority of studies reported data with HAQ-DI and SF-36 PCS, while only four and one reported data with SF-36 PF and HAQ-S, respectively. This systematic review supports clinical trial discrimination with HAQ-DI and SF-36 PCS with low risk of bias, whereas clinical trial discrimination with SF-36 PF and HAQ-S are supported with caution. There are no data published to date to support clinical trial discrimination for potential PF-PROMs shortlisted by GRAPPA such as MDHAQ and PROMIS.

Physical function is one of the domains included in the core outcome set for reporting data in PsA RCTs and longitudinal observational studies (6). GRAPPA and OMERACT are committed to standardizing outcome measures based on evidence but variations in reporting outcomes in PsA clinical trials have been recognized (11). To appropriately evaluate the measurement properties of instruments using the OMERACT Filter 2.1, multiple measurement properties are considered, including domain match, feasibility, validity, test-retest reliability, longitudinal construct validity, clinical trial discrimination and threshold of meaning (15). In this study, we aimed to evaluate only the clinical trial discrimination of PF-PROMs. Although this represents an intermediate step in standardizing the outcome measurement set for physical function in PsA, it is important in the process.

The strength of the current work is the combined effort of investigators and patient research partners (PRPs). The investigators are familiar with the measurement of physical function in PsA with representation from 4 continents while the PRPs have participated in a wide range of research activities, including data extraction, quality assessment of articles and appraisal of ESs. We followed the methods recommended by the OMERACT Filter 2.1 methodology in quality assessment of each article, calculating ESs using appropriate statistics to synthesize the evidence to support clinical trial discrimination (15). This was further strengthened by setting *a priori* hypotheses of the expected magnitude of ESs of interventional compared with control groups.

Some limitations are recognized. We limited the evidence to RCTs with b-DMARDs and ts-DMARDs; and excluded RCTs evaluating solely cs-DMARDs. We think this best represents RCTs in the modern era that include the appropriate core domains. We calculated the ESs of PF-PROMs from published data instead of using the original dataset from the RCTs. Where applicable, we used formulas estimating means (SD) from reported medians (IQR), which may result in variability. Nonetheless, this variability in ES estimations has been recognized and addressed in the quality assessment using the OMERACT good method checklist; and the detailed calculations are shown in the supplementary documents. Estimated ESs (ES_1, ES_2 or SRM) were tabulated depending on availability of published data, which may not be directly comparable. This would contribute only to minor differences as all hypothesis testing was performed intra- rather than inter- trials using the same type of ESs.

CONCLUSION

This systematic review supports clinical trial discrimination with HAQ-DI and SF-36 PCS with low risk of bias, while clinical trial discrimination with SF-36 PF and HAQ-S were supported with some caution.

List of Abbreviations

ABT: abatacept;
ADA: adalimumab;
ALC: alefacept;
b-DMARDs: biological disease modifying anti-rheumatic drugs;
BIW: twice a week;
BRO: brodalumab;
CI: confidence interval;
CLZ: clazakizumab;
cs-DMARDs: conventional synthetic disease modifying anti-rheumatic drugs;
CZP: certolizumab pegol;
ETN: etanercept;
EULAR: European League Against Rheumatism;
FIL: filgotinib;
GOL: golimumab;
GRAPPA: Group for Research and Assessment of Psoriasis and Psoriatic Arthritis
GUS: guselkumab;
HAQ-DI: Health Assessment Questionnaire – Disability Index;
HAQ-S: Health Assessment Questionnaire – Spine;
IFX: infliximab;
IL: interleukin;
IQR: interquartile range;
IV: intravenous
IXE: ixekizumab;
LS: least squares;
MCID: minimally clinically important difference;
MTX: methotrexate;
NA: not available;
NS: not significant;
OMERACT: Outcome Measures in Rheumatology
PASI: Psoriasis Area and Severity Index;
PCB: placebo;
PCS: physical component summary of SF-36;
PF: physical functioning domain of SF-36;
PROM: patient reported outcome measure;
PsA: psoriatic arthritis;
Q2W: once every 2 weeks;
QW: once a week;
RCTs: randomized controlled trials
SD: standard deviation;
SE: standard error;
SEC: secukinumab;
SF-36: Medical Outcomes Study 36-item Short Form Survey;
TOF: tofacitinib;
TNFi: tumor necrosis factor inhibitor;
ts-DMARDs: targeted synthetic disease modifying anti-rheumatic drugs;
UST: ustekinumab;
vs.: versus.

References

1. Gladman DD, Antoni C, Mease P, Clegg DO, Nash P. Psoriatic arthritis: epidemiology, clinical features, course, and outcome. Annals of the rheumatic diseases. 2005;64 Suppl 2:ii14-7.
2. Gladman DD, Farewell VT, Nadeau C. Clinical indicators of progression in psoriatic arthritis: multivariate relative risk model. The Journal of rheumatology. 1995;22(4):675-9.
3. Husted JA, Tom BD, Farewell VT, Schentag CT, Gladman DD. Description and prediction of physical functional disability in psoriatic arthritis: a longitudinal analysis using a Markov model approach. Arthritis and rheumatism. 2005;53(3):404-9.
4. Orbai AM, de Wit M, Mease PJ, Callis Duffin K, Elmamoun M, Tillett W, et al. Updating the Psoriatic Arthritis (PsA) Core Domain Set: A Report from the PsA Workshop at OMERACT 2016. The Journal of rheumatology. 2017;44(10):1522-8.
5. Dures E, Hewlett S, Lord J, Bowen C, McHugh N, Group PS, et al. Important Treatment Outcomes for Patients with Psoriatic Arthritis: A Multisite Qualitative Study. Patient. 2017;10(4):455-62.
6. Orbai AM, de Wit M, Mease P, Shea JA, Gossec L, Leung YY, et al. International patient and physician consensus on a psoriatic arthritis core outcome set for clinical trials. Ann Rheum Dis. 2017;76(4):673-80.
7. Mease P, Strand V, Gladman D. Functional impairment measurement in psoriatic arthritis: Importance and challenges. Semin Arthritis Rheum. 2018;48(3):436-48.
8. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. Arthritis Rheum. 1980;23(2):137-45.
9. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care. 1992;30(6):473-83.
10. Hojgaard P, Klokker L, Orbai AM, Holmsted K, Bartels EM, Leung YY, et al. A systematic review of measurement properties of patient reported outcome measures in psoriatic arthritis: A GRAPPA-OMERACT initiative. Seminars in arthritis and rheumatism. 2018;47(5):654-65.
11. Tillett W, Orbai AM, Ogdie A, Leung YY, Strand V, Gladman DD, et al. GRAPPA-OMERACT initiative to standardise outcomes in psoriatic arthritis clinical trials and longitudinal observational studies. Ann Rheum Dis. 2018;77(5):e23.
12. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ (Clinical research ed). 2009;339:b2700.
13. Collaboration TC. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0. In: Higgins JPT, Green S, editors. 2011.
14. Wan X, Wang W, Liu J, Tong T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. BMC Med Res Methodol. 2014;14:135.
15. Beaton DE, Maxwell LJ, Shea BJ, Wells GA, Boers M, Grosskleg S, et al. Instrument Selection Using the OMERACT Filter 2.1: The OMERACT Methodology. J Rheumatol. 2019;46(8):1028-35.
16. Daltroy LH, Larson MG, Roberts NW, Liang MH. A modification of the Health Assessment Questionnaire for the spondyloarthropathies. J Rheumatol. 1990;17(7):946-50.
17. Kavanaugh A, McInnes IB, Mease PJ, Hall S, Chinoy H, Kivitz AJ, et al. Efficacy of Subcutaneous Secukinumab in Patients with Active Psoriatic Arthritis Stratified by Prior Tumor Necrosis Factor Inhibitor Use: Results from the Randomized Placebo-controlled FUTURE 2 Study. The Journal of rheumatology. 2016;43(9):1713-7.
18. Ritchlin C, Rahman P, Kavanaugh A, McInnes IB, Puig L, Li S, et al. Efficacy and safety of the anti-IL-12/23 p40 monoclonal antibody, ustekinumab, in patients with active psoriatic arthritis despite conventional non-biological and biological anti-tumour necrosis factor therapy: 6-month and

1-year results of the phase 3, multicentre, double-blind, placebo-controlled, randomised PSUMMIT 2 trial. Annals of the rheumatic diseases. 2014;73(6):990-9.
19. Mease PJ, Gottlieb AB, van der Heijde D, FitzGerald O, Johnsen A, Nys M, et al. Efficacy and safety of abatacept, a T-cell modulator, in a randomised, double-blind, placebo-controlled, phase III study in psoriatic arthritis. Annals of the rheumatic diseases. 2017;76(9):1550-8.
20. Mease PJ, Genovese MC, Weinblatt ME, Peloso PM, Chen K, Othman AA, et al. Phase II Study of ABT-122, a Tumor Necrosis Factor– and Interleukin-17A–Targeted Dual Variable Domain Immunoglobulin, in Patients With Psoriatic Arthritis With an Inadequate Response to Methotrexate. Arthritis and Rheumatology. 2018;70(11):1778-89.
21. Mease PJ, Genovese MC, Greenwald MW, Ritchlin CT, Beaulieu AD, Deodhar A, et al. Brodalumab, an anti-IL17RA monoclonal antibody, in psoriatic arthritis. The New England journal of medicine. 2014;370(24):2295-306.
22. Antoni CE, Kavanaugh A, Kirkham B, Tutuncu Z, Burmester GR, Schneider U, et al. Sustained benefits of infliximab therapy for dermatologic and articular manifestations of psoriatic arthritis: results from the infliximab multinational psoriatic arthritis controlled trial (IMPACT). Arthritis and rheumatism. 2005;52(4):1227-36.
23. Antoni C, Krueger GG, de Vlam K, Birbara C, Beutler A, Guzzo C, et al. Infliximab improves signs and symptoms of psoriatic arthritis: results of the IMPACT 2 trial. Annals of the rheumatic diseases. 2005;64(8):1150-7.
24. Kavanaugh A, Antoni C, Krueger GG, Yan S, Bala M, Dooley LT, et al. Infliximab improves health related quality of life and physical function in patients with psoriatic arthritis. Annals of the rheumatic diseases. 2006;65(4):471-7.
25. Mease PJ, Ory P, Sharp JT, Ritchlin CT, Van den Bosch F, Wellborne F, et al. Adalimumab for long-term treatment of psoriatic arthritis: 2-year data from the Adalimumab Effectiveness in Psoriatic Arthritis Trial (ADEPT). Annals of the rheumatic diseases. 2009;68(5):702-9.
26. Genovese MC, Mease PJ, Thomson GT, Kivitz AJ, Perdok RJ, Weinberg MA, et al. Safety and efficacy of adalimumab in treatment of patients with psoriatic arthritis who had failed disease modifying antirheumatic drug therapy. The Journal of rheumatology. 2007;34(5):1040-50.
27. Mease PJ, Goffe BS, Metz J, VanderStoep A, Finck B, Burge DJ. Etanercept in the treatment of psoriatic arthritis and psoriasis: a randomised trial. Lancet (London, England). 2000;356(9227):385-90.
28. Mease PJ, Woolley JM, Singh A, Tsuji W, Dunn M, Chiou CF. Patient-reported outcomes in a randomized trial of etanercept in psoriatic arthritis. The Journal of rheumatology. 2010;37(6):1221-7.
29. Gniadecki R, Robertson D, Molta CT, Freundlich B, Pedersen R, Li W, et al. Self-reported health outcomes in patients with psoriasis and psoriatic arthritis randomized to two etanercept regimens. Journal of the European Academy of Dermatology and Venereology : JEADV. 2012;26(11):1436-43.
30. Mease PJ, Gladman DD, Collier DH, Ritchlin CT, Helliwell PS, Liu L, et al. Etanercept and Methotrexate as Monotherapy or in Combination for Psoriatic Arthritis: Primary Results From a Randomized, Controlled Phase 3 Trial. Arthritis & rheumatology (Hoboken, NJ). 2019.
31. Kavanaugh A, McInnes I, Mease P, Krueger GG, Gladman D, Gomez-Reino J, et al. Golimumab, a new human tumor necrosis factor alpha antibody, administered every four weeks as a subcutaneous injection in psoriatic arthritis: Twenty-four-week efficacy and safety results of a randomized, placebo-controlled study. Arthritis and rheumatism. 2009;60(4):976-86.
32. Kavanaugh A, Husni ME, Harrison DD, Kim L, Lo KH, Leu JH, et al. Safety and Efficacy of Intravenous Golimumab in Patients With Active Psoriatic Arthritis: Results Through Week Twenty-Four of the GO-VIBRANT Study. Arthritis & rheumatology (Hoboken, NJ). 2017;69(11):2151-61.
33. Gladman D, Fleischmann R, Coteur G, Woltering F, Mease PJ. Effect of certolizumab pegol on multiple facets of psoriatic arthritis as reported by patients: 24-week patient-reported outcome results of a phase III, multicenter study. Arthritis care & research. 2014;66(7):1085-92.

34. McInnes IB, Sieper J, Braun J, Emery P, van der Heijde D, Isaacs JD, et al. Efficacy and safety of secukinumab, a fully human anti-interleukin-17A monoclonal antibody, in patients with moderate-to-severe psoriatic arthritis: a 24-week, randomised, double-blind, placebo-controlled, phase II proof-of-concept trial. Annals of the rheumatic diseases. 2014;73(2):349-56.
35. Mease PJ, McInnes IB, Kirkham B, Kavanaugh A, Rahman P, van der Heijde D, et al. Secukinumab Inhibition of Interleukin-17A in Patients with Psoriatic Arthritis. The New England journal of medicine. 2015;373(14):1329-39.
36. McInnes IB, Mease PJ, Kirkham B, Kavanaugh A, Ritchlin CT, Rahman P, et al. Secukinumab, a human anti-interleukin-17A monoclonal antibody, in patients with psoriatic arthritis (FUTURE 2): a randomised, double-blind, placebo-controlled, phase 3 trial. Lancet (London, England). 2015;386(9999):1137-46.
37. Nash P, Mease PJ, McInnes IB, Rahman P, Ritchlin CT, Blanco R, et al. Efficacy and safety of secukinumab administration by autoinjector in patients with psoriatic arthritis: Results from a randomized, placebo-controlled trial (FUTURE 3). Arthritis Research and Therapy. 2018;20(1).
38. Mease PJ, van der Heijde D, Ritchlin CT, Okada M, Cuchacovich RS, Shuler CL, et al. Ixekizumab, an interleukin-17A specific monoclonal antibody, for the treatment of biologic-naive patients with active psoriatic arthritis: results from the 24-week randomised, double-blind, placebo-controlled and active (adalimumab)-controlled period of the phase III trial SPIRIT-P1. Annals of the rheumatic diseases. 2017;76(1):79-87.
39. Nash P, Kirkham B, Okada M, Rahman P, Combe B, Burmester GR, et al. Ixekizumab for the treatment of patients with active psoriatic arthritis and an inadequate response to tumour necrosis factor inhibitors: results from the 24-week randomised, double-blind, placebo-controlled period of the SPIRIT-P2 phase 3 trial. Lancet (London, England). 2017;389(10086):2317-27.
40. Gottlieb A, Menter A, Mendelsohn A, Shen YK, Li S, Guzzo C, et al. Ustekinumab, a human interleukin 12/23 monoclonal antibody, for psoriatic arthritis: randomised, double-blind, placebo-controlled, crossover trial. Lancet (London, England). 2009;373(9664):633-40.
41. McInnes IB, Kavanaugh A, Gottlieb AB, Puig L, Rahman P, Ritchlin C, et al. Efficacy and safety of ustekinumab in patients with active psoriatic arthritis: 1 year results of the phase 3, multicentre, double-blind, placebo-controlled PSUMMIT 1 trial. Lancet (London, England). 2013;382(9894):780-9.
42. Araujo EG, Englbrecht M, Hoepken S, Finzel S, Kampylafka E, Kleyer A, et al. Effects of ustekinumab versus tumor necrosis factor inhibition on enthesitis: Results from the enthesial clearance in psoriatic arthritis (ECLIPSA) study. Seminars in arthritis and rheumatism. 2019;48(4):632-7.
43. Deodhar A, Gottlieb AB, Boehncke WH, Dong B, Wang Y, Zhuang Y, et al. Efficacy and safety of guselkumab in patients with active psoriatic arthritis: a randomised, double-blind, placebo-controlled, phase 2 study. Lancet (London, England). 2018;391(10136):2213-24.
44. Mease P, Genovese MC, Gladstein G, Kivitz AJ, Ritchlin C, Tak PP, et al. Abatacept in the treatment of patients with psoriatic arthritis: results of a six-month, multicenter, randomized, double-blind, placebo-controlled, phase II trial. Arthritis and rheumatism. 2011;63(4):939-48.
45. Mease PJ, Gladman DD, Keystone EC. Alefacept in combination with methotrexate for the treatment of psoriatic arthritis: results of a randomized, double-blind, placebo-controlled study. Arthritis and rheumatism. 2006;54(5):1638-45.
46. Mease P, Hall S, FitzGerald O, van der Heijde D, Merola JF, Avila-Zapata F, et al. Tofacitinib or Adalimumab versus Placebo for Psoriatic Arthritis. The New England journal of medicine. 2017;377(16):1537-50.
47. Gladman D, Rigby W, Azevedo VF, Behrens F, Blanco R, Kaszuba A, et al. Tofacitinib for Psoriatic Arthritis in Patients with an Inadequate Response to TNF Inhibitors. The New England journal of medicine. 2017;377(16):1525-36.
48. Mease P, Coates LC, Helliwell PS, Stanislavchuk M, Rychlewska-Hanczewska A, Dudek A, et al. Efficacy and safety of filgotinib, a selective Janus kinase 1 inhibitor, in patients with active psoriatic

arthritis (EQUATOR): results from a randomised, placebo-controlled, phase 2 trial. Lancet (London, England). 2018;392(10162):2367-77.
49. Mease PJ, Gottlieb AB, Berman A, Drescher E, Xing J, Wong R, et al. The Efficacy and Safety of Clazakizumab, an Anti-Interleukin-6 Monoclonal Antibody, in a Phase IIb Study of Adults With Active Psoriatic Arthritis. Arthritis & rheumatology (Hoboken, NJ). 2016;68(9):2163-73.
50. Mease PJ, Gladman DD, Ritchlin CT, Ruderman EM, Steinfeld SD, Choy EH, et al. Adalimumab for the treatment of patients with moderately to severely active psoriatic arthritis: results of a double-blind, randomized, placebo-controlled trial. Arthritis and rheumatism. 2005;52(10):3279-89.

Figure 1. Flow diagram for article selection

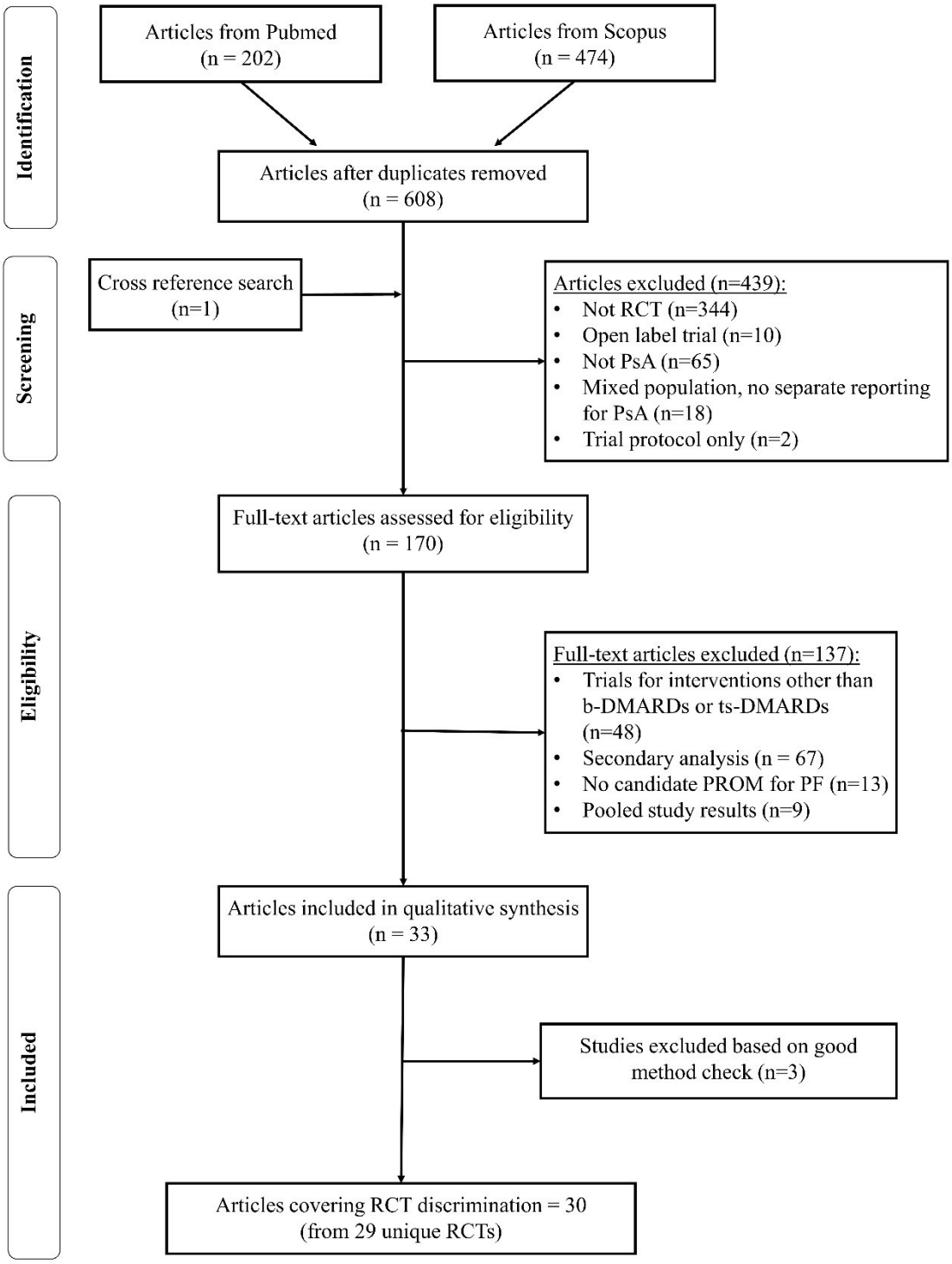

Table 1. Characteristics of included studies

Author/ year/ (study acronyms)	Intervention/ Comparator	Sample size (% women)	PsA duration (years)	TNFi IR	MTX use	Primary outcome/ time point	Baseline HAQ-DI (mean)	Reported data for PF-PROM			
								HAQ-DI	HAQ-S	SF-36 PCS	SF-36 PF
TNF inhibitors											
Antoni, et al. 2005 (IMPACT) (22)	IFX 5mg/kg vs. PCB	N=104 (42.3)	11.4	0%	NA	ACR20/ Week 16	1.2	Yes	No	No	No
Antoni, et al. 2005 (IMPACT2) (23)	IFX 5mg/kg vs. PCB	N=200 (39.0)	8	0%	46%	ACR20/ Week 14	1.1	Yes	No	-	No
Kavanaugh, et al. 2006 (IMPACT2) (24)								-	-	Yes	-
Mease, et al. 2005 (ADEPT) (25)	ADA 40mg Q2W vs. PCB	N=313 (44.4)	9.5	0%	51%	ACR20/ Week 12 Δ in modified total sharp score/ week 24	1.0	Yes	No	Yes	No
Genovese, et al. 2007 (26)	ADA 40mg Q2W vs. PCB	N=100 (46.0)	7.4	0%	47% (cs-DAMRD 66%)	ACR20/ Week 12	1.0	Yes	No	Yes	No
Mease, et al. 2000 (27)	ETN 25mg BIW vs. PCB	N=60 (43.0)	9.3	0%	47%	PsARC/ Week 12	1.3	Yes	No	No	No
Mease, et al. 2010 (28)	ETN 25mg BIW vs. PCB	N=205 (49.0)	9.1	0%	41.5%	ACR20/ Week 12	1.1	Yes	No	Yes	No
Gniadecki, et al. 2012 (PRESTA) (29)	ETN 50mg BIW/QW vs. ETN 50mg QW/QW	N=752 (37.0)	7.0	NA	NA	Psoriasis clear or almost clear/ Week 12	0.92	Yes	No	No	No
Mease, et al. 2019 (SEAM-PsA) (30)	ETN 50mg QW vs. ETN 50mg QW plus MTX vs. MTX alone	N=851 (50.8)	3.2	0%	100% (prior MTX use 0%)	ACR20/ Week 24	1.2	Yes	No	Yes	No
Kavanaugh, et al. 2009 (GO-REVEAL) (31)	GOL (2 doses) vs. PCB	N=405 (39.8)	7.5	0%	48%	ACR20/ Week 14	1.3	Yes	No	Yes	No
Kavanaugh, et al.	GOL IV 2mg/kg	N=480	5.8	0%	70%	ACR20/	1.3	Yes	No	Yes	No

Author/ year/ (study acronyms)	Intervention/ Comparator	Sample size (% women)	PsA duration (years)	TNFi IR	MTX use	Primary outcome/ time point	Baseline HAQ-DI (mean)	Reported data for PF-PROM			
								HAQ-DI	HAQ-S	SF-36 PCS	SF-36 PF
2017 (GO-VIBRANT) (32)	vs. PCB	(48.0)				Week 14					
Gladman, et al. 2014 (RAPID-PsA) (33)	CZP pegol (2 doses) vs. PCB	N=409 (55.3)	8.6	19.1%	64.1%;	ACR20, EULAR response/ Week 12	1.3	Yes	No	Yes	No
IL17 inhibitors											
McInnes, 2014 (Phase II) (34)	SEC (10mg/kg) vs. PCB (N=42)	N=42 (64.0)	NA	35%	49% (any cs-DMARD 51%)	ACR20/ Week 6	1.5	Yes	No	Yes	No
Mease, et al. 2015 (FUTURE I) (35)	SEC (2 doses) vs PCB (N=606)	N=606 (54.5)	NA	29.4%	60.7%;	ACR20/ Week 24*	1.2	Yes	No	Yes	No
McInnes, et al. 2015 (FUTURE II) (36)	SEC (3 doses) vs. PCB	N=397 (51.6)	NA	35%	47%;	ACR20/ Week 24*	1.2	Yes	No	Yes	No
Kavanaugh, et al. 2016. (FUTURE II) (17) -subgroup analysis	TNFi naïve vs. TNFi exposed										
Nash, et al. 2018 (FUTURE III) (37)	SEC (2 doses) vs. PCB	N=414 (54.8)	7.5	32%	47.6%	ACR20/ Week 24	1.2	Yes	No	Yes	No
Mease, et al. 2017 (SPIRIT-P1) (38)	IXE (2 doses) vs. PCB vs. ADA	N=417 (54.0)	6.7	0%	54.2% (any cs-DMARD 64%)	ACR20/ Week 24*	1.2	Yes	No	Yes	No
Nash, et al. 2017 (SPIRIT-P2) (39)	IXE (2 doses) vs. PCB	N=363 (53.4)	10.0	100%	41%	ACR20/ Week 24*	1.2	Yes	No	Yes	No
Mease, et al. 2014 Phase II (21)	BRO (2 doses) vs. PCB	N=168 (64.0)	8.7	51%	50%	ACR20/ Week 12	1.3	Yes	No	Yes	No
IL12/23 inhibitors											
Gottlieb, et al. 2009 (40)	UST (2 doses) vs. PCB	N=146 (43.8)	5.6	27.4%	20.5%	ACR20/ Week 12	0.9	Yes	No	No	No
McInnes, et al. 2013 (PSUMMIT I) (41)	UST (2 doses) vs. PCB	N=615 (46.3)	4.0	0%	48%	ACR20, EULAR response PASI75/	1.3	Yes	No	Yes	No

Author/ year/ (study acronyms)	Intervention/ Comparator	Sample size (% women)	PsA duration (years)	TNFi IR	MTX use	Primary outcome/ time point	Baseline HAQ-DI (mean)	Reported data for PF-PROM			
								HAQ-DI	HAQ-S	SF-36 PCS	SF-36 PF
						Week 24*					
Ritchlin, et al. 2014 (PSUMMIT II) (18)	UST (2 doses) vs. PCB	N=312 (52.6)	5.1	57.7%	49.7%;	ACR20, EULAR response PASI75/ Week 24*	1.0	Yes	No	Yes	No
Araugo, et al. 2019 (ECLIPSA) (42)	UST (45mg or 100mg if body weight >100kg) vs. TNFi	N=47 (40.4)	2.5	0%	91.5%	Enthesitis clearance (SPARCC=0)/ Week 24	1.0	Yes	No	Yes	No
IL23 inhibitors											
Deodhar, et al. 2018 (43)	GUS 100mg vs. PCB	N=149 (49.0)	7.0	8.7%	44.3%	ACR20/ Week24*	1.4	Yes	No	Yes	No
T cell inhibition											
Mease, et al. 2011 Phase II (44)	ABT (3 doses) vs. PCB	N=170 (44.2)	8.2	37.4%	58.3%	ACR20/ Day 169	1.2	Yes	No	Yes	No
Mease, et al. 2017 (ASTRAEA) (19)	ABT vs. PCB	N=424 (45.0)	8.5	61%	60%	ACR20/ Week 24*	1.3	Yes	No	Yes	No
Mease, et al. 2006 (45)	ALC/MTX vs. PCB/MTX	N=185 (61.1)	5	0%	100%	ACR20/ Week 12	1.1	Yes	No	No	No
JAK inhibitors											
Mease, et al. 2017 (OPAL Broaden) (46)	TOF (2 doses) vs. PCB vs. ADA	N=422 (53.0)	6.1	0%	83.9%	ACR20/ 3-month	1.1	Yes	No	Yes	No
Gladman, et al. 2017 (OPAL Beyond) (47)	TOF (2 doses) vs. PCB	N=395 (55.0)	9.4	100%	73.6%	ACR20, Δ in HAQ-DI/ 3-month	1.3	Yes	No	No	Yes
Mease, et al. 2018 (EQUATOR) (48) Phase II	FIL 200mg vs. PCB	N=131 (50.4)	7.0	0%	54.2% (any cs-DMARD 74%)	ACR20/ Week 16	1.4	Yes	No	No	Yes
IL6 inhibitors											
Mease, et al. 2016 Phase II (49)	CLZ (3 doses) vs. PCB	N=165 (59.4)	7.1	0%	69.1%	ACR20/ Week 16	1.4	Yes	No	Yes	No

Author/ year/ (study acronyms)	Intervention/ Comparator	Sample size (% women)	PsA duration (years)	TNFi IR	MTX use	Primary outcome/ time point	Baseline HAQ-DI (mean)	Reported data for PF-PROM			
								HAQ-DI	HAQ-S	SF-36 PCS	SF-36 PF
Others											
Mease, et al. 2018 Phase II (20)	ABT-122 (2 doses) vs. PCB vs. ADA	N=240 (49.6)	7.3	0%	100%	ACR20/ Week 12	1.3	No	Yes	No	No

*early escape at Week 16.

Abbreviations: Δ: change; ACR: American College of Rheumatology Response criteria; ABT: abatacept; ADA: adalimumab; ALC: alefacept; BIW: twice a week; BRO: brodalumab; CI: confidence interval; CLZ: clazakizumab; cs-DMARDs: conventional synthetic disease modifying anti-rheumatic drugs; CZP: certolizumab pegol; ETN: etanercept; EULAR: European League Against Rheumatism; FIL: filgotinib; GOL: golimumab; GUS: guselkumab; HAQ-DI: Health Assessment Questionnaire – Disability Index; IFX: infliximab; IL=interleukin; IQR: interquartile range; IV: intravenous; IXE: ixekizumab; LS: least squares; MCID: minimally clinically important difference; MTX: methotrexate; NA: not available; NS: not significant; PsARC: Psoriatic Arthritis Response Criteria; PASI: Psoriasis Area and Severity Index; PCB: placebo; PCS: physical component summary of SF-36; PF: physical functioning domain of SF-36; QW: once a week; Q2W: once every 2-week; SD: standard deviation; SE: standard error; SEC: secukinumab; SF-36: Medical Outcomes Study 36-item Short Form Survey; SPARCC: Spondyloarthritis Research Consortium of Canada enthesitis index; TNFi: tumor necrosis factor inhibitors; TOF: tofacitinib; UST: ustekinumab; vs.: versus.

Table 2. Effect size estimation for studies reporting HAQ-DI

Author/ year/ (study acronyms)	Intervention/ comparator (sample size, N)	Primary outcome/ time point	Effect sizes at primary endpoint (unless specified)	Fulfillment of *a priori* hypothesis
Antoni, et al. 2005 (IMPACT) (22)	IFX 5mg/kg vs. PCB (N=104)	ACR20/ Week 16	SRM[¶] (for improvement)[δ] at Week 16: IFX = 6.07 PCB = -0.19	1, 2, 3
Antoni, et al. 2005 (IMPACT2) (23)	IFX 5mg/kg vs. PCB (N=200)	ACR20/ Week 14	SRM[¶] (for improvement)[δ] at Week 14: IFX = 1.08 PCB = -0.19	1, 2, 3
Mease, et al. 2005 (ADEPT) (50)	ADA 40mg Q2W vs. PCB (N=313)	ACR20/ Week 12; Δ in modified Total Sharp Score/ Week 24	SRM at Week 12: ADA = -0.8 PCB = 0.2	1, 2, 3
Genovese, et al. 2007 (26)	ADA 40mg Q2W vs. PCB (N=100)	ACR20/ Week 12	SRM at Week 12: ADA = -0.6 PCB = -0.33	1, 2, 3
Mease, et al. 2000 (27)	ETN 25mg BIW vs. PCB (N=60)	PsARC/ Week 12	ES_2[¥] at Week 12: ETN = 0.547 PCB = 0.237	1, 2, 3
Mease, et al. 2010 (28)	ETN 25mg BIW vs. PCB (N=205)	ACR20/ Week 12 Week 24 end of double-blind phase	ES for Week 12: NA ES_2 at Week 24 (end of double-blind phase): ETN = -0.597 PCB = -0.098	1, 2, 3
Gniadecki, et al. 2012 (PRESTA) (29)	ETN 50mg BIW/QW vs. ETN 50mg QW/QW (N=752)	Psoriasis clear or almost clear/ Week 12	ES_2 at Week 12: ETN 50mg BIW/QW: -0.74 ETN 50mg QW/QW: -0.69	1, 3
Mease, et al. 2019 (SEAM-PsA) (30)	ETN 50mg QW vs. ETN 50mg QW plus MTX vs. MTX alone (N=851)	ACR20/ Week 24	SRM at Week 24: ETN = -0.733 ETN plus MTX = -0.685 MTX alone = -0.646	1, 3
Kavanaugh, et al. 2009 (GO-REVEAL) (31)	GOL (2 doses) vs PCB (N=405)	ACR20/ Week 14	SRM at Week 14: GOL 100mg = -0.75 GOL 50mg = -0.62 PCB = -0.09	1, 2, 3

Author/ year/ (study acronyms)	**Intervention/ comparator (sample size, N)**	**Primary outcome/ time point**	**Effect sizes at primary endpoint (unless specified)**	**Fulfillment of *a priori* hypothesis**
Kavanaugh, et al. 2017 (GO-VIBRANT) (32)	GOL IV 2mg/kg vs. PCB (N=480)	ACR20/ Week 14	SRM at Week 14: GOL IV = -1.13 PCB = -0.26	1, 2, 3
Gladman, et al. 2014 (RAPID-PsA) (33)	CZP (2 doses) vs. PCB (N=409)	ACR20, EULAR response/ Week 12	SRM at Week 12: CZP 400mg Q4W = -0.83 CZP 200mg Q2W = -0.80 PCB = -0.44	1, 2, 3
McInnes, 2014 (Phase II) (34)	SEC (10mg/kg) vs. PCB (N=42)	ACR20/ Week 6	SRM at Week 6: SEC = -0.680 PCB = 0.018	1, 2, 3
Mease, et al. 2015 (FUTURE I) (35)	SEC (2 doses with loading) vs. PCB (N=606)	ACR20/ Week 24*	SRM at Week 24: SEC 150mg = -0.703 SEC 75mg = -0.721 PCB = 0.239	1, 2, 3
McInnes, et al. 2015 FUTURE II) (36)	SEC (3 doses) vs. PCB (N=397)	ACR20/ Week 24*	SRM at Week 24: SEC 300mg = -1.12 SEC 150mg = -0.96 SEC 75mg = -0.644 PCB = -0.522	1, 2, 3
Kavanaugh, et al. 2016 (FUTURE II) (17) -subgroup analysis	TNFi naïve vs. TNFi exposed	ACR20/ Week 24*	SRM at Week 24 (TNFi naïve vs. exposed): SEC 300mg: -1.20 vs. -1.02 SEC 150mg: -1.15 vs. -0.71 SEC 75mg: -0.77 vs. -0.44 PCB: -0.63 vs -0.35	4
Nash, et al. 2018 (FUTURE III) (37)	SEC (2 doses) vs. PCB (N=414)	ACR20/ Week 24*	SRM at Week 24: SEC 300mg = -0.81 SEC 150mg = -0.57 PCB = -0.24	1, 2, 3
Mease, et al. 2017 (SPIRIT-P1) (38)	IXE (2 doses) vs. PCB vs. ADA (N=417)	ACR20/ Week 24*	SRM at Week 24: IXE Q2W = -0.98 IXE Q4W = -0.85 PCB = -0.35 ADA = -0.74	1, 2, 3
Nash, et al. 2017 (SPIRIT-P2) (39)	IXE (2 doses) vs. PCB (N=363)	ACR20/ Week 24*	SRM at Week 24: IXE Q2W = -0.36 IXE Q4W = -0.55 PCB = -0.18	1, 2, 3

Author/ year/ (study acronyms)	**Intervention/ comparator (sample size, N)**	**Primary outcome/ time point**	**Effect sizes at primary endpoint (unless specified)**	**Fulfillment of *a priori* hypothesis**
Mease, et al. 2014 Phase II (21)	BRO (2 doses) vs. PCB (N=168)	ACR20/ Week 12	SRM[¥] at Week 12: BRO 280mg = -0.60 BRO 140mg = -0.38 PCB = -0.21	1, 2, 3
Gottlieb, et al. 2009 (40)	UST (2 doses) vs. PCB (N=146)	ACR20/ Week 12	SRM[¥] at Week 12: UST = -0.66 PCB = -0.14	1, 2, 3
McInnes, et al. 2013 (PSUMMIT I) (41)	UST (2 doses) vs. PCB (N=615)	ACR20, EULAR response, PASI75/ Week 24*	SRM[¥] at Week 24: UST 90mg = -0.59 UST 45mg = -0.62 PCB = -0.21	1, 2, 3
Ritchlin, et al. 2014 (PSUMMIT II) (18)	UST (2 doses) vs. PCB (N=312)	ACR20, EULAR response, PASI75/ Week 24*	SRM[¥] at Week 24: UST 90mg = -0.66 UST 45mg = -0.59 PCB = 0.00 SRM[¥] at Week 24: Subgroup analysis: TNFi naïve vs. exposed UST 90mg = -0.66 vs. -0.66 UST 45mg = -0.66 vs. -0.59 PCB = 0 vs. 0	1, 2, 3
Araugo, et al. 2019 (ECLIPSA) (42)	UST (45mg or 100mg if body weight >100kg) vs. TNFi (N=47)	Enthesitis clearance (SPARCC = 0)/ Week 24	ES_2[¥] at Week 24: UST = 1.81 TNFi = 1.74	1, 2, 3
Deodhar, et al. 2018 (43)	GUS 100mg vs. PCB (N=149)	ACR20/ Week 24*	SRM at Week 24: GUS = -0.82 PCB = -0.11	1, 2, 3
Mease, et al. 2011 Phase II (44)	ABT (3 doses) vs. PCB (N=170)	Day 169	Insufficient data for ES calculation	none

Author/ year/ (study acronyms)	**Intervention/ comparator (sample size, N)**	**Primary outcome/ time point**	**Effect sizes at primary endpoint (unless specified)**	**Fulfillment of *a priori* hypothesis**
Mease, et al. 2017 (ASTRAEA) (19)	ABT vs. PCB (N=424)	ACR20/ Week 24*	SRM[¥] at Week 24: ABT = 0.69 PCB = 0.40 SRM[¥] at Week 24: Subgroup analysis: TNFi naïve vs. exposed ABT = 0.62 vs. 0.64 PCB = 0.39 vs. 0.34	1, 3
Mease, et al. 2006 (45)	ALC/MTX vs. PCB/MTX (N=185)	ACR20/ Week 12	% Δ in HAQ-DI: ALC/MTX = -24.5% PCB/MTX = -7.7% (NS) Inadequate data for ES calculations	none
Mease, et al. 2017 (OPAL Broaden) (46)	TOF (2 doses) vs. PCB vs. ADA (N= 422)	ACR20/ 3 months	SRM at 3 months: TOF 10mg = -0.78 TOF 5mg = -0.69 PCB = -0.36 ADA = -0.76	1, 2, 3
Gladman, et al. 2017 (OPAL Beyond) (47)	TOF (2 doses) vs. PCB (N=395)	ACR20, Δ in HAQ-DI/ 3 months	SRM at 3 months: TOF 10mg = -0.64 TOF = -0.70 PCB = -0.26	1, 2, 3
Mease, et al. 2018 (EQUATOR) (48) Phase II	FIL 200mg vs. PCB (N=131)	ACR20/ Week 16	SRM at Week 16: FIL = 1.14 PCB = 0.56 ES_2 at Week 16: FIL = 1.04 PCB = 0.47	1, 2, 3
Mease, et al. 2016 Phase II (49)	CLZ (3 doses) vs. PCB (N=165)	ACR20/ Week 16	SRM[¥] at Week 16: CLZ 200mg = -0.51 CLZ 100mg = -0.77 CLZ 50mg = -0.83 PCB = -0.52	1, 2, 3

[¶] SRM calculated using percentage change score and SD of percentage change; [δ] SRM for improvement, a negative value indicate deterioration; [¥] Effect sizes estimated based on mean and SD of change scores calculated from median and IQR from original publication; *early escape for patients with inadequate response in the control group to active treatment group at Week 16; **option to switch TNFi at Week 24.

Abbreviations: Δ: change; ACR: American College of Rheumatology Response criteria; ABT: abatacept; ADA: adalimumab; ALC: alefacept; BIW: twice a week; BRO: brodalumab; CI: confidence interval; CLZ: clazakizumab; CZP: certolizumab pegol; ES_2: Effect size 2 (the mean difference divided by the pooled standard deviation, i.e., Cohen's *d*); ETN: etanercept; EULAR: European League Against Rheumatism; FIL: filgotinib; GOL: golimumab; GUS: guselkumab; HAQ-DI: Health Assessment Questionnaire – Disability Index; IFX: infliximab; IQR: interquartile range; IV: intravenous; IXE: ixekizumab; LS: least squares; MCID: minimally clinically important difference; MTX: methotrexate; NA: not available; NS: not significant; PsARC: Psoriatic Arthritis Response Criteria; PASI: Psoriasis Area and Severity Index; PCB: placebo; PCS: physical component summary of SF-36; PF: physical functioning domain of SF-36; QW: once a week; Q2W: once every 2 weeks; SD: standard deviation; SE: standard error; SEC: secukinumab; SF-36: Medical Outcomes Study 36-item Short Form Survey 36 items; SPARCC: Spondyloarthritis Research Consortium of Canada enthesitis index; SRM: Standardized response mean (mean difference divided by the standard deviation of the differences between baseline and assessment end point); TNFi: tumor necrosis factor inhibitors; TOF: tofacitinib; UST: ustekinumab; vs.: versus.

Table 3. Effect size estimation for studies reporting HAQ-S

Author/ year/ (study acronyms)	Intervention/ comparator (sample size, N)	Primary outcome/ time point	PROM	Effect sizes at primary end point (unless specified)	Fulfillment of *a priori* hypothesis
Mease, et al. 2018 Phase II (20)	ABT-122 (2 doses) vs. PCB vs. ADA (N=240)	ACR20/ Week 12	HAQ-S	ES_1 at Week 12: ABT-122 240mg: -0.93 ABT-122 120mg: -0.92 PCB: -0.47 ADA: -0.97	3

Abbreviations: ACR: American College of Rheumatology Response criteria; ADA: adalimumab; ES_1: Effect size 1 (the mean difference divided by standard deviation of baseline score; HAQ-S: Health Assessment Questionnaire – Spine; PCB: placebo;

Table 4. Effect size estimation for studies reporting SF-36 PCS

Author / year/ (study acronyms)	Intervention/ comparator (sample size, N)	Primary outcome/ time point	Effect sizes at primary end point (unless specified)	Fulfillment of *a priori* hypothesis
Antoni, et al. 2005 (IMPACT2) (23)	IFX 5mg/kg vs. PCB (N=200)	ACR20/ Week 14	SRM at Week 14: IFX = 0.98 PCB = 0.13	1, 2, 3
Mease, et al. 2005 (ADEPT) (50)	ADA 40mg Q2W vs. PCB (N=313)	ACR20/ Week 12; Δ in modified Total Sharp Score/ Week 24	SRM at Week 12: ADA = 0.93 PCB = 0.16	1, 2, 3
Genovese, et al. 2007 (26)	ADA 40mg Q2W vs. PCB (N=100)	ACR20/ Week 12	SRM at Week 12: ADA = 0.67 PCB = 0.39	1, 2, 3
Mease, et al. 2010 (28)	ETN 25mg BIW vs. PCB (N=205)	ACR20/ Week 12 Week 24 end of double-blind phase	SRM at Week 12: NA ES_2 at Week 24 (end of double-blind phase): ETN = 0.880 PCB = 0.073	1, 2, 3
Mease, et al. 2019 (SEAM-PsA) (30)	ETN 50mg QW vs. ETN 50mg QW + MTX vs. MTX alone (N=851)	ACR20/ Week 24	SRM at Week 24: ETN = 0.832 ETN plus MTX = 0.813 MTX alone = 0.629	1, 2, 3
Kavanaugh, et al. 2009 (GO-REVEAL) (31)	GOL (2 doses) vs PCB (N=405)	ACR20/ Week 14	SRM at Week 14: GOL 100mg = 0.82 GOL 50mg = 0.74 PCB = 0.08	1, 2, 3
Kavanaugh, et al. 2017 (GO-VIBRANT) (32)	GOL IV 2mg/kg vs. PCB (N=480)	ACR20/ Week 14	SRM at Week 14: GOL IV = 1.14 PCB = 0.46	1, 2, 3
Gladman, et al. 2014 (RAPID-PsA) (33)	CZP (2 doses) vs. PCB (N=409)	ACR20, EULAR response/ Week 12	SRM at Week 12: CZP 400mg Q4W = 0.87 CZP 200mg Q2W = 0.82 PCB = 0.30	1, 2, 3
McInnes IB, 2014 (Phase II) (34)	SEC (10mg/kg) vs. PCB (N=42)	ACR20/ Week 6	SRM at Week 6: SEC = 0.541 PCB = -0.017	1, 2, 3

Author / year/ (study acronyms)	**Intervention/ comparator (sample size, N)**	**Primary outcome/ time point**	**Effect sizes at primary end point (unless specified)**	**Fulfillment of *a priori* hypothesis**
Mease, et al. 2015 (FUTURE I) (35)	SEC (2 doses with loading) vs. PCB (N=606)	ACR20/ Week 24*	SRM at Week 24: SEC 150mg = 0.785 SEC 75mg = 0.732 PCB = 0.178	1, 2, 3
McInnes, et al. 2015 FUTURE II) (36)	SEC (3 doses) vs. PCB (N=397)	ACR20/ Week 24*	SRM at Week 24: SEC 300mg = 0.98 SEC 150mg = 0.875 SEC 75mg = 0.587 PCB = 0.203	1, 2, 3
Kavanaugh, et al. 2016. (FUTURE II) (17) -subgroup analysis	TNFi naïve vs. TNFi exposed	ACR20/ week 24*	SRM at Week 24: (TNFi naïve vs. exposed): SEC 300mg: 1.07 vs. 0.95 SEC 150mg: 1.07 vs. 0.60 SEC 75mg: 0.71 vs. 0.45 PCB: 0.22 vs 0.27	4
Nash, et al. 2018 (FUTURE III) (37)	SEC (2 doses) vs. PCB (N=414)	ACR20/ Week 24*	SRM at Week 24: SEC 300mg = 0.93 SEC 150mg = 0.49 PCB = 0.30	1, 2, 3
Mease, et al. 2017 (SPIRIT-P1) (38)	IXE (2 doses) vs. PCB vs. ADA (N=417)	ACR20/ Week 24*	SRM at Week 24: IXE Q2W = 1.79 IXE Q4W = 1.33 PCB = 0.55 ADA = 1.78	1, 2, 3
Nash, et al. 2017 (SPIRIT-P2) (39)	IXE (2 doses) vs. PCB (N=363)	ACR20/ Week 24*	SRM at Week 24: IXE Q2W = 0.30 IXE Q4W = 0.27 PCB = 0.06	1, 2, 3
Mease, et al. 2014 Phase II (21)	BRO (2 doses) vs. PCB (N=168)	ACR20/ Week 12	Difference from PCB (95% CI): BRO 280mg: 2.4 (0.1 to 4.6) BRO 140mg: 1.4 (-0.8 to 3.6) Insufficient data for ES calculation	none
McInnes, et al. 2013 (PSUMMIT I) (41)	UST (2 doses) vs. PCB (N=615)	ACR20, EULAR response, PASI75 /Week 24*	SRM[¥] at Week 24: UST 90mg = 0.75 UST 45mg = 0.48 PCB = 0.24	1, 2, 3

Author / year/ (study acronyms)	Intervention/ comparator (sample size, N)	Primary outcome/ time point	Effect sizes at primary end point (unless specified)	Fulfillment of *a priori* hypothesis
Ritchlin, et al. 2014 (PSUMMIT II) (18)	UST (2 doses) vs. PCB (N=312)	ACR20, EULAR response, PASI75/ Week 24*	SRM[¥] at Week 24: UST 90mg = 0.60 UST 45mg = 0.50 PCB = 0.30	1, 2, 3
Araugo, et al. 2019 (ECLIPSA) (42)	UST (45mg or 100mg if body weight >100kg) vs. TNFi (N=47)	Enthesitis clearance (SPARCC = 0)/ Week 24	ES_2[¥] at Week 24: UST = 2.95 TNFi = 1.56	1, 3
Deodhar, et al. 2018 (43)	GUS 100mg vs. PCB (N=149)	ACR20/ Week 24*	SRM at Week 24: GUS = 0.88 PCB = 0.06	1, 2, 3
Mease, et al. 2011 Phase II (44)	ABT (3 doses) vs. PCB (N=170)	Day 169	SRM at Day 169: ABT 30/10 mg/kg: 0.59 ABT 10mg/kg: 0.77 ABT 3 mg/kg: 0.53 PCB: 0.02	1, 2, 3
Mease, et al. 2017 (ASTRAEA) (19)	ABT vs. PCB (N=424)	ACR20/ Week 24*	SRM[¥] at Week 24: ABT: 0.72 PCB: 0.53	1, 2, 3
Mease, et al. 2016 Phase II (49)	CLZ (3 doses) vs. PCB (N=165)	ACR20/ Week 16	SRM[¥] at Week 16: CLZ 200mg = 0.52 CLZ 100mg = 0.59 CLZ 50mg = 0.82 PCB = 0.57	1, 2, 3

[¶] SRM calculated using percentage change score and SD of percentage change; [δ] SRM for improvement, a negative value indicate deterioration; [¥] Effect sizes estimated based on mean and SD of change scores calculated from median and IQR from original publication; * early escape for patients with inadequate response in the PCB group to active treatment group at Week 16; ** option to switch TNFi at Week 24;

Abbreviations: Δ: change; ACR: American College of Rheumatology Response criteria; ABT: abatacept; ADA: adalimumab; ALC: alefacept; BIW: twice a week; BRO: brodalumab; CI: confidence interval; CLZ: clazakizumab; CZP: certolizumab; ES_2: Effect size 2 (the mean difference divided by the pooled standard deviation, i.e., Cohen's *d*); ETN: etanercept; EULAR: European League Against Rheumatic Diseases; FIL: filgotinib; GOL: golimumab; GUS: guselkumab; HAQ-DI: Health Assessment Questionnaire – Disability index; IFX: infliximab; IQR: interquartile range; IV: intravenous; IXE: ixekizumab; LS: least squares; MCID: minimally clinically important difference; MTX: methotrexate; NA: not available; NS: not significant; PsARC: Psoriatic Arthritis Response Criteria; PASI: Psoriasis Area and Severity Index; PCB: placebo; PCS: physical component summary of SF-36; PF: physical functioning domain of SF-36; QW: once a week; Q2W: once every 2 weeks; SD: standard deviation; SE: standard error; SEC: secukinumab; SF-36: Medical Outcomes Study 36-item Short Form Survey; SPARCC: Spondyloarthritis Research Consortium of Canada enthesitis index; SRM: Standardized response mean (mean difference divided by the standard deviation of the differences between baseline and assessment end point); TNFi: tumor necrosis factor inhibitors; TOF: tofacitinib; UST: ustekinumab; vs.: versus.

Table 5. Effect size estimation of all studies reporting SF-36 PF

Author/ year/ (study acronyms)	Intervention/ comparator (sample size, N)	Primary outcome/ time point	Effect sizes at primary endpoint (unless specified)	Fulfillment of *a priori* hypothesis
Antoni, et al. 2007 (IMPACT2) (23)	IFX 5mg/kg vs. PCB (N=200)	ACR20/ Week 14	Insufficient data for effect size calculation	1, 2
Mease, et al. 2017 (OPAL Broaden)	TOF (2 doses) vs. PCB vs. ADA (N= 422)	ACR20/ 3 months	SRM at 3 months: TOF 10mg = 0.64 TOF 5mg = 0.64 PCB = 0.23 ADA = 0.58	1, 3
Gladman, et al. 2017 (OPAL Beyond)	TOF (2 doses) vs. PCB (N=395)	ACR20, Δ in HAQ-DI/ 3 months	SRM at 3 months: TOF 10mg = 0.53 TOF = 0.64 PCB = 0.22	1, 3
Mease, et al. 2018 (EQUATOR) Phase II	FIL 200mg vs. PCB (N=131)	ACR20/ Week 16	Insufficient data for effect size calculation	1, 2

[¶] SRM calculated using percentage change score and SD of percentage change; [δ] SRM for improvement, a negative value indicate deterioration; [¥] Effect sizes estimated based on mean and SD of change scores calculated from median and IQR from original publication; * early escape for patients with inadequate response in the control group to active treatment group at Week 16; ** option to switch TNFi at Week 24;
Abbreviations: Δ: change; ACR: American College of Rheumatology Response criteria; ABT: abatacept; ADA: adalimumab; ALC: alefacept; BIW: twice a week; BRO: brodalumab; CI: confidence interval; CLZ: clazakizumab; CZP: certolizumab pegol; ES_2: Effect size 2 (the mean difference divided by the pooled standard deviation, i.e Cohen's *d*); ETN: etanercept; EULAR: European League Against Rheumatism; FIL: filgotinib; GOL: golimumab; GUS: guselkumab; HAQ-DI: Health Assessment Questionnaire – Disability index; IFX: infliximab; IQR: interquartile range; IV: intravenous; IXE: ixekizumab; LS: least squares; MCID: minimally clinically important difference; MTX: methotrexate; NA: not available; NS: not significant; PsARC: Psoriatic Arthritis Response Criteria; PASI: Psoriasis Area and Severity Index; PCB: placebo; PCS: physical component summary of SF-36; PF: physical functioning domain of SF-36; QW: once a week; Q2W: once every 2 weeks; SD: standard deviation; SE: standard error; SEC: secukinumab; SF-36: Medical Outcomes Study 36-item Short Form Survey; SPARCC: Spondyloarthritis Research Consortium of Canada enthesitis index; SRM: Standardized response mean (mean difference divided by the standard deviation of the differences between baseline and assessment end point); TNFi: tumor necrosis factor inhibitors; TOF: tofacitinib; UST: ustekinumab; vs.: versus.

Table 6. Summary of Measurement Properties Table for clinical trial discrimination

1st Author/ year/ (study acronyms/ drug)	HAQ-DI	HAQ-S	SF-36 PCS	SF-36 PF
Antoni, 2005 (IMPACT)	+			
Antoni, 2005 (IMPACT2)	+		+	
Kavanaugh, 2006 (IMPACT2)				+
Mease, 2005 (ADEPT)	+		+	
Genovese, 2007 (ADA)	+		+	
Mease, 2000 (ETN)	+		+	
Mease, 2010 (ETN)	+		+	
Gniadecki, 2012 (PRESTA)	+			
Mease. 2019 (SEAM-PsA)	+/-		+	
Kavanaugh, 2009 (GO-REVEAL)	+		+	
Kavanaugh, 2017 (GO-VIBRANT)	+		+	
Gladman, 2014 (RAPID-PsA)	+		+	
McInnes, 2014 (SEC)	+		+	
Mease, 2015 (FUTURE I)	+		+	
McInnes, 2015 (FUTURE II)	+		+	
Kavanaugh, 2016 (FUTURE II) -subgroup analysis	+		+	
Nash P, 2018 (FUTURE III)	+		+	
Mease, 2017 (SPIRIT-P1)	+		+	
Nash, 2017 (SPIRIT-P2)	+		+	
Mease, 2014 (BRO)	+		+	
Gottlieb, 2009 (UST)	+			
McInnes, 2013 (PSUMMIT I)	+		+	
Ritchlin, 2014 (PSUMMIT II)	+		+	
Araugo, 2019 (ECLIPSA)	+		+	
Deodhar, 2018 (GUS)	+		+	
Mease, 2011 (ABT)	+/-		+	
Mease, 2017 (ASTRAEA)	+		+	
Mease, 2006 (ALC)	+/-			
Mease, 2017 (OPAL Broaden)	+			+
Gladman, 2017 (OPAL Beyond)	+			+
Mease, 2018 (EQUATOR)	+			+
Mease, 2016 (CLZ)	+		+	
Mease, 2018 (ABT-122)		+		
Total available studies	**31**	**1**	**24**	**4**
Total studies for evidence synthesis	**29**	**1**	**23**	**2**
Overall rating	+	+	+	+

Color code in each box indicates study quality assessed by OMERACT good methods. GREEN means "yes, likely low risk of bias"; AMBER means "some cautions but can be used as evidence" and RED means "No, don't use as evidence". WHITE (empty boxes), indicates absence of information on that property from that study. (+) indicates findings of the study had adequate performance of the instrument; (+/-) indicates equivocal performance; (-) indicates poor performance (less than adequate).